Review

The Role of Constitutional Copy Number Variants in Breast Cancer

Logan C. Walker [1],*, George A.R. Wiggins [1], John F. Pearson [2]

[1] Mackenzie Cancer Research Group, Department of Pathology, University of Otago, Christchurch 8140, New Zealand; E-Mail: george.wiggins@otago.ac.nz

[2] Biostatistics and Computational Biology Unit, University of Otago, Christchurch 8140, New Zealand; E-Mail: john.pearson@otago.ac.nz

* Author to whom correspondence should be addressed; E-Mail: logan.walker@otago.ac.nz; Tel.: +64-3-364-0544.

Academic Editor: Massimo Negrini

Abstract: Constitutional copy number variants (CNVs) include inherited and *de novo* deviations from a diploid state at a defined genomic region. These variants contribute significantly to genetic variation and disease in humans, including breast cancer susceptibility. Identification of genetic risk factors for breast cancer in recent years has been dominated by the use of genome-wide technologies, such as single nucleotide polymorphism (SNP)-arrays, with a significant focus on single nucleotide variants. To date, these large datasets have been underutilised for generating genome-wide CNV profiles despite offering a massive resource for assessing the contribution of these structural variants to breast cancer risk. Technical challenges remain in determining the location and distribution of CNVs across the human genome due to the accuracy of computational prediction algorithms and resolution of the array data. Moreover, better methods are required for interpreting the functional effect of newly discovered CNVs. In this review, we explore current and future application of SNP array technology to assess rare and common CNVs in association with breast cancer risk in humans.

1. Introduction

Over the past decade there have been a large number of studies that have explored the biological impact of constitutional (inherited and *de novo*) copy number variants (CNVs) in the human genome [1,2]. CNVs are structural rearrangements that increase or decrease DNA content at regions larger than 50 base pairs (bps) in size [1,2], accounting for a majority of genetic variation in humans based on bp coverage. These variants are estimated to cover 5%–10% [2] of the human genome which is at least an order of magnitude greater than the number of bps (~15 Mbps; dbSNP Human Build 142) encompassed by the more commonly studied single nucleotide polymorphisms (SNPs).

Molecular technologies used to profile DNA copy number, such as microarrays (SNP-based arrays and comparative genomic hybridisation) and next-generation sequencing, have led to the identification of more than 300,000 CNVs, or 21,757 unique CNV loci in the human genome [3] . These technologies have also revealed the extent to which constitutional CNVs partially overlap or fully encompass genes and/or regulatory sequences. Concomitant gene expression analyses have shown a strong relationship between copy number dosage and mRNA levels with hundreds of genes [4,5]. This functional effect can play an important role in a variety of human diseases, including breast cancer [6–9].

2. Single Nucleotide Polymorphism (SNP)-Array Platforms to Assess Breast Cancer Risk

A significant proportion of breast cancers arise in a subset of women who have multiple affected relatives as a result of inherited genetic factors that increase the risk of developing the disease. The relative risk (RR) of breast cancer in mothers and sisters of patients is increased, ranging from 1.8-fold to more than 5-fold [10,11]. In 5%–10% patients, inherited mutations in highly penetrant cancer susceptibility genes, such as *BRCA1* and *BRCA2*, are known to confer a significantly elevated risk (>10-fold) of breast cancer and their carrier relatives [12]. A further 5% of cases carry deleterious variants in moderate-risk breast cancer susceptibility genes, such as *CHEK2*, *ATM*, *BRIP1*, and *PALB2* [11–14]. However, these variants are too rare to be identified in most genome-wide association studies and do not increase risk sufficiently for capture by linkage analysis in family studies.

Numerous genome-wide association studies for different population groups have successfully been performed to discover low-risk SNP variants that are associated with breast cancer [15–33]. Such studies have been underpinned by SNP array platforms from companies, such as Affymetrix, Illumina and Perlegen Sciences, ranging in genome coverage, spatial resolution and design. Probes used on SNP arrays for these studies have generally been selected to target SNPs with a minor allele frequency greater than 5%. Thus, genome-wide association studies are designed to detect causal variants that are relatively common in the population. As breast cancer studies have grown in size, less common variants are able to be assessed for risk association. A recent initiative as part of the Collaborative Oncological Gene-Environment Study (COGS) used a custom-designed array to assess almost 200,000 SNPs across the genome in approximately 50,000 breast cancer cases and 50,000 controls [28]. Studies of this size are statistically powered to evaluate variants with a minor allele frequency <5%. As a result of the large COGS initiative, more than 90 independent common susceptibility loci have now been identified, explaining a further 16% of the familial risk [27].

Currently known low-, moderate- and high-risk genetic factors explain up to half of the familial clustering in breast cancer [28]; thus, for a substantial fraction of women, the genetic changes

contributing to breast cancer remains undetermined, even if they have a family history [34]. Discovery of variants to explain this "missing heritability" is of clinical relevance, but will require different approaches that perhaps include other types of genetic variation, such as CNVs, using high throughput technology.

3. Copy Number Variant (CNV) Prediction Algorithms for SNP Array Data

The ability to study CNVs at a genome-wide level has been made possible by the development of high-throughput SNP array technologies. Moreover, the vast amount of SNP-genotyping data generated by numerous genome-wide association studies of breast cancer offers significant potential to explore the contribution of CNVs to this disease. SNP markers present on many early Affymetrix and Illumina arrays were also supplemented with thousands of intensity-only (non-polymorphic) probes that target known CNV regions, especially those regions unsuitable for SNP genotyping probes.

A large number of CNV calling algorithms have been applied to SNP array and/or array comparative hybridisation data in published studies with variable success. A proportion of these algorithms have been utilised more frequently for a variety of reasons, including accuracy, availability and suitability to the array platform used in the studies and ease of implementation. Most algorithms are either proprietary and available commercially, or have coded implementations freely available for downloading. Table 1 lists those in common use by the citations of their principal publication in PubMed at the time of writing. A measure we acknowledge underestimates the popularity of commercial (and usually unpublished) solutions.

ACCURACY of CNV Predictions from SNP Arrays

A major limitation for the use of SNP arrays in CNV association studies is the accuracy of CNV calling algorithms. The current CNV algorithms vary in methodology and subsequently produce varied results (Table 1). The most numerous CNV calling methods use Hidden Markov Models (HMM) to estimate copy number at loci with transition probabilities estimated or supplied, as for example from gold standard datasets. Others methods use mixtures—particularly Gaussian—distributions, or Bayesian methods. Many implementations include heuristics to deal with or explicitly model features in the data such as loss of heterozygosity regions and GC waves, and set a minimum number of probes for which they will call a CNV.

Methods have been proposed that might reduce false positives, including altering parameters within the algorithms (*e.g.* CNV size and number of probes included) and comparing multiple algorithms [35]. Validation of predicted structural variants is critical for the use in association studies. Table 2 provides a list of studies that explored the issue of algorithm accuracy. Three studies [36–38] assessed the accuracy of multiple CNV calling algorithms by comparing data they derived from samples previously used in "gold standard" studies [39,40]. These reports present different conclusions with respect to algorithm performance, although PennCNV was the only algorithm included in all three studies. Winchester and colleagues validated 49% of CNVs predicted by PennCNV in the Kidd *et al.* [40] study for the highest rate in their study. Zhang and colleagues used multiple permutations to obtain the greatest recovery of CNVs from gold standard studies using the same samples. For PennCNV with pedigree information included, a maximum recovery rate (number of CNVs in Conrad *et al.* [39] that were

predicted) was only 35% using >20 probes. Birdsuite was able to recover nearly half of the predicted CNVs (48%) under similar setting (no pedigree information). Zhang *et al* [38] found deletions were validated at a much higher rate with both Partek and Birdsuite correctly predicting deletions selected for validation (5/5). In comparison, predicted duplications showed a high false positive rate with PennCNV, the most accurate predicting 66.7% of CNVs validated (4/6) [38]. Similarly, Seiser and Innocenti assessed three samples previously characterised in Conrad *et al.* [39] to measure the performance of three HMM algorithms (GenoCN, PennCNV and QuantiSNP) [36]. PennCNV performed poorly with low sensitivity (14.46%, minimum of five probes) and high specificity (a common trait for HMM algorithms). With exception of Zhang *et al.* [38], many studies were limited by the reliance on CNVs from previously published reports as there was no attempt to experimentally validate predicted variants. Zhang and colleagues illustrate this vulnerability by highlighting disagreement with commonly used gold standards from Conrad *et al.* [39] and Kidd *et al.* [40]. Comparing CNVs calls in five samples used by each study showing strikingly poor agreement [38]. Other studies have used mass spectrometry, quantitative polymerase chain reaction (qPCR) and/or multiplex ligation-dependent probe amplification (MLPA) to attempt to validate CNVs [41,42]. Typically, these studies used methods to reduce false positives by creating strict criteria for inclusion. One study confirmed that sensitivity was a weakness of CNVPartition, PennCNV and QuantiSNP, with QuantiSNP showing the greatest MLPA-validated sensitivity (28%) [42]. This study also showed that, of the true positives, each algorithm tended to correctly predict the CNV class (homozygous deletion, heterozygous deletion and duplication) with sensitivity >92% and specificity >87%. An exception to these results was the ability of QuantiSNP to accurately call homozygous and heterozygous deletions, with call rates of 68% and 62%, respectively). Together, these studies highlight the lack of a consensus on CNV-calling methodologies used to assess SNP array data. Furthermore, results from publications reviewed in Table 1 support the necessity to experimentally validate any CNV loci that are predicted by SNP array data, and are to be included in breast cancer association studies

Table 1. Commonly (>10 citations) applied CNV detection methods for SNP-array data.

Software	Algorithm	Code	Platform	Year [a]	Reference	Citations [b]	Software URL
PennCNV	HMM	Perl	Multiple	2007	[43]	300	http://penncnv.openbioinformatics.org
Birdsuite (Birdseye, Canary)	Mixture models	Java/Python/R	Affymetrix	2008	[44]	300	http://www.broadinstitute.org
Nexus Copy Number	Proprietary (Segmentation)	windows executable	Multiple	-	-	100	http://www.biodiscovery.com
QuantiSNP	HMM	MATLAB	Multiple	2007	[45]	100	http://sites.google.com/site/quantisnp
CNVPartition	Proprietary	windows executable	Illumina	2006	-	100	http://support.illumina.com
Partek Genomics Suite	Proprietary (Segmentation or HMM)	windows executable	Multiple	-	-	30	http://www.partek.com/pgs
CNVFinder	Experimental variability	perl	Array CGH	2006	[46]	30	http://www.sanger.ac.uk/resources/software/cnvfinder/
CGHCall	segmentation and mixture model	R	Array CGH	2007	[47]	30	http://www.few.vu.nl/~mavdwiel/CGHcall.html
GenoCNV	HMM	R	Multiple	2009	[48]	30	http://www.bios.unc.edu/~weisun/software/genoCN.htm
SW-ARRAY	Smith Waterman	R	Array CGH	2005	[49]	30	Not available
HMMSeg	HMM wavelet smoothing	Java	Multiple	2007	[50]	10	http://noble.gs.washington.edu/proj/hmmseg
VanillaICE	HMM	R	Affymetrix	2008	[51]	10	http://cran.r-project.org
CNVHap	HMM, Haplotype	Java	Multiple	2010	[52]	10	http://www.imperial.ac.uk/people/l.coin
dChip	Multiple	R	Multiple	2008	[53]	10	http://sites.google.com/site/dchipsoft
GADA	Bayesian	R	Multiple	2010	[54]	10	http://cran.r-project.org
CNV Workshop	Segmentation	complete VM	Multiple	2010	[55]	10	http://sourceforge.net/projects/cnv

[a] Year reference when published. [b] At least this many citations in PubMed or company website at July 2015. Abbreviation: HMM, Hidden Markov Model.

Table 2. Accuracy of CNV-calling algorithms.

Algorithm(s)	Platform	Validation Method	Accuracy	Study Conclusion	Reference
Adapted method on SW-ARRAY and GIM	Affymetrix	qPCR or Mass Spec Validation	2.5% false positives, ~90% singleton validation	Developed a multistep algorithm to better call CNVs.	[41]
Birdsuite, CNAT, CNVPartition, GADA, Nexus, PennCNV and QuantiSNP	Affymetrix, Illumina	Comparison of HapMap samples to Kidd *et al.*, Korbel *et al.* and Redon *et al.*, data [5,40,56]	Assay sensitivity ranged 20%−49% with some algorithms predicting more events (*i.e.*, GADA, 546 predicted CNVs).	PennCNV had the greatest sensitivity (49%). Little agreement between studies and within studies.	[37]
cnvHap, CNVPartition, PennCNV and QuantiSNP	Aglient, Illumnina	Compared samples either with previously characterized (by aCGH) CNVs or HapMap samples from Kidd *et al.* [40]	cnvHap had very good sensitivity (68%) for larger CNVs (>10kb) in Kidd *et al.* This reduced to 31% for smaller CNVs (<5kb).	cnvHap has increased sensitivity compared with other CNV algorithms.	[52]
PennCNV, Aroma.Affymetrix, APT and CRLMM	Affymetrix	Compared concordance between calling algorithms.	Greater concordance in deletion (51.5%) than duplications (47.9%). The probable false positive rates for CRLMM and PennCNV were 26% and 24%.	PennCNV appeared to detect all the CNV and more than CRLMM predicted	[57]
CNVPartition, PennCNV and QuantiSNP	Illumnina	Agreement between algorithms	Agreement varied from 59%−62% for deletions, to 43%−57% for duplications.	Use of multiple algorithms increased the positive predictive value, as did the number of probes and the minimum size (kb).	[35]
CNVPartition, PennCNV and QuantiSNP	Illumnina	MLPA validation, measures were taken to reduce false positive calls.	All algorithms show better specificity than sensitivity. QuantiSNP was the most sensitive, predicting 28% of CNVs. PennCNV was better at discriminating copy number state.	Applying methods to reduce false positives results in low sensitivity.	[42]
ADM-2, Birdsuite, CNVfinder, CNVPartition, dCHIP, GTC, iPattern, Nexus, Partek, PennCNV, QuantiSNP	CGH arrays and SNP arrays (Affymetrix and Illumina)	Experiments were repeated in triplicate and CNV calls were compared. CNV calls were also compared to 5 references ('gold standards').	Algorithm replication has <70% reproducibility. CNV calls between any two algorithms is typically low (25%–50%) within a platform. Overlap with DGV was high, whereas overlap with references [39,40] was low.	Newer high resolution arrays outperform older arrays in both CNVs' call and reproducibility. Algorithms developed for specific array platforms outperformed adapted and independent algorithms.	[58]

Table 2. *Cont*.

Algorithm(s)	Platform	Validation Method	Accuracy	Study Conclusion	Reference
Birdsuite, Partek, Genomics Suite, HelixTree and PennCNV	Affymetrix	Comparison with HapMap CNV in two studies [39,40]. qPCR validation of rare CNVs (a single CNV event in >1000 bipolar samples)	Overlap ranged between 42% and 70% when including 20 probes for Kidd *et al.* [40] and 26%−48% in Conrad *et al.* [39] For each algorithm between 10 or 11, CNVs were tested. Partek and Birdsuite both validated all (5/5) deletion events tested.	Birdsuite outperformed the other 3 algorithms over multiple permutation. Birduite and Partek had high positive predictive values, particularly with deletions. HelixTree performed poorly.	[38]
CNVPartition, PennCNV and QuantiSNP	Illumnina	Comparison to a previous CGH study [59]. qPCR validation of 3 candidate loci in 717 horses.	50 CNVs were called by all 3 algorithms. QuantiSNP had the highest overlap with CNVs predicted from CGH arrays (25%). Validation rates were greater than 80% for the 3 loci.	CNVPartition predicted the least CNVs, suggesting a high false negative rate.	[60]
GenoCN, PennCNV and QuantiSNP	Illumnina	Comparison of HapMap sample to Conrad *et al.*[39] Compared both CNVs (*i.e.* Gain or Loss) and normal calls.	All algorithms show much better specificity than sensitivity. PennCNV had the worst sensitivity, predicting <15% of Conrad *et al.* [39] CNVs in 3 samples	The three HMM algorithms all performed with varied results. They were all highly specific (>98%), but sensitivity remains to be an issue for all three algorithms.	[36]
cnvHap, COKGEN, GenoCNV, HaplotypeCN, PennCNV and QuantiSNP	Affymetrix	Compared 270 HapMap samples which have been previously described. Compared simulated data to test haplotype phasing between cnvHap and HaplotypeCNV.	GenoCNV has the most sensitivity (28%) when using Kidd *et al.* [40]; however, the concordance rate in PennCNV was greater (36% and 9%, respectively).	Algorithm performance varied with reference study. GenoCNV was the most sensitive but had the lowest concordance rate. HaplotypeCNV, cnvHap and PennCNV (under a specific permutation) were compared separately, with HaplotypeCN outperforming the other two.	[61]
Birdsuite, dCHIP, GTC and PennCNV	Affymetrix	Comparison to a previous CGH study [62].	GTC had the highest portion of CNV matching (50% overlap) to CGH, 66%. Larger CNVs were called with greater accuracy.	Birdsuite called the most CNVs; however, PennCNV outperformed all algorithms with greater specificity and sensitivity.	[63]

Abbreviations: aCGH, array comparative genomic hybridisation; APT, Affymetrix Power Tools; CNV, copy number variant; CRLMM, corrected robust linear mixture model; DGV, Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/home); HMM, hidden Markov model; GTC, Genotyping Console; kb, kilobases; MLPA, Multiplex ligation-dependent probe amplification; qPCR, quantitative polymerase chain reaction.

4. Functional Annotation of CNVs

The functional impact of CNVs in the human genome vary as a result of the variant size, copy number state, and location relative to genes or key regulatory regions. Homozygous deletions overlapping at least 85% of exons from approximately 100 protein-coding genes have been identified in genomes from seemingly healthy individuals [2], suggesting these genes are functionally redundant or are related to an unknown phenotype. Haploinsufficiency for genes disrupted by a hemizygous deletion is also an important mechanism for genetic disease, such as *APOBEC3B* and breast cancer risk [6,64]. Conversely, gene duplications resulting from overlapping CNVs can influence biology through triplosensitivity.

There is an increasing number of CNVs of unknown clinical significance that are predicted to be involved in disease susceptibility due to potentially deleterious effects on overlapping or nearby gene(s). Despite the myriad of computational tools developed to detect CNVs for different array and sequencing platforms, a significant informatics challenge exists for interpreting both the functional and clinical role of these variants. Computational tools, such as SG-ADVISER CNV [65], CNV-WebStore [66] and CNVannotator [67], have been developed to derive functional effects from predicted variants. These tools are useful for assigning potential clinical implications of CNVs based on their location within known pathogenic regions. To assess variant pathogenicity, SG-ADVISER CNV utilises additional factors to generate a classification score, including 1) allele frequency information from repositories, such as the 1000 Genomes Project; and 2) clinical genetic information from databases, such as Online Mendelian Inheritance in Man [68], ClinVar [69]. However, a major limitation of annotating CNV regions derived using SNP arrays is the inability to precisely define their breakpoints. Thus, any overlap between predicted CNVs with clinically relevant regions along the genome remain putative without further validation using ancillary techniques, such as quantitative PCR or MLPA.

5. Application of SNP Arrays for Profiling CNVs in Breast Cancer

Structural variants, including CNVs, contribute to many complex diseases, and could account for some of the missing heritability of breast cancer. CNVs have been reported to encompass genes known to be involved in breast cancer susceptibility, including *BRCA1* and *BRCA2,* and therefore may similarly affect other genes involved in breast cancer-related pathways [12].

5.1. Inherited Copy Number Polymorphisms and Breast Cancer Risk

Analysis of large genome-wide association studies carried out by the Wellcome Trust Case Control Consortium suggested that common CNVs were unlikely to play a major role in breast cancer susceptibility [70]. This study used a 105K probe Agilent CGH array design containing probes tagging for copy number loci previously identified from (1) the Genome Structural Variation (GSV) Consortium [39]; (2) CNV studies using the SNP arrays Affymetrix 6.0, Illumina 1M, and Affymetrix 500k; (3) novel sequence absent from the reference sequence; 4) candidate genes; and 5) additional risk-associated loci. However, this study was not sufficiently powered to detect the effects of low-penetrant alleles with a minor allele frequency (MAF) less than 5%. Moreover, the genomic regions assessed by this study were limited by the design of the arrays used to generate genotype information across the genome. More recently, a genome-wide association study of common CNVs (MAF $\geq$ 5%) conducted among Chinese

women using high-resolution data from the Affymetrix SNP Array 6.0 identified a deletion in the *APOBEC3* gene cluster associated with breast cancer risk. Within this population, the deletion was identified in 65% cases and 45% of controls, conferring odds ratios (ORs) of 1.3 and 1.8 for a hemizygous and homozygous deletion, respectively ($p = 2.0 \times 10^{-24}$) [6]. Subsequent investigations of women with European ancestry using quantitative-PCR also observed the deletion, albeit at a much lower population frequency [71]. Comparable to the study of Chinese women, a higher proportion of breast cancer affected European women (12.4% *vs.* 10.4%, respectively) because they carried the *APOBEC3* allele, thereby conferring low to moderate risk of disease (ORs of 1.2 and 2.3 ($p = 0.005$) for a hemizygous and homozygous deletion, respectively). Interestingly, the same deletion (CNV ID: CNVR8164.1) was originally identified by the Wellcome Trust Case Control Consortium; however, replication experiments did not show a significant association with breast cancer.

As mentioned above, there is now a wealth of array data available from SNP-based genome-wide association studies that can be utilised for assessing the contribution of CNVs to breast cancer risk. Furthermore, the huge number of cases and controls available for future CNV association studies will provide sufficient power to evaluate many CNVs that occur at low frequency. A major limitation with using these array data is the inability to genotype highly repetitive copy number-variable regions. More than 1000 regions across the human genome have been found overlapping CNVs with three or more segregating alleles [72]. Non-array-based technologies that can resolve multicopy integer states, such as qPCR, Nanostring and massively parallel sequencing, will therefore be necessary to determine the clinical significance of these multiallelic variants in breast cancer and other human diseases.

5.2. Inherited and de novo Rare CNVs and Breast Cancer Risk

At least seven array-based studies have reported lists of rare CNVs overlapping genes that may contribute towards the development breast cancer [8,73,74]. Despite a number of candidate susceptibility genes being proposed there has been a notable lack of concordance between these studies. More than 120 genes overlapping rare genomic deletions or duplications have been found exclusively or at a greater frequency in familial breast cancer cases; however, none have been replicated between studies (Supplementary Table S1). Such a finding is not surprising as many individuals carry rare or private CNVs regardless of their disease status [2,75]. Furthermore, four of these studies used SNP-based arrays which are known to generate signal-to-noise ratios that are much lower than array-CGH platforms and are therefore more prone to false CNV calls [58]. It remains unclear whether future large-scale studies will provide the reproducible evidence needed to implicate these rare CNVs as breast cancer risk variants and to overcome the issue of false discovery.

Growing evidence suggests that the frequency and size of constitutional CNVs are significantly increased in breast cancer-affected individuals [73,74,76]. Studies have assessed the global burden of deletions and duplications in cases and controls by measuring: (1) the number of CNVs per sample; (2) the number CNVs overlapping genes (and *vice versa*) per sample; (3) the average length of CNVs per sample; and (4) the total number of base pairs affected by CNVs per sample. Although studies have revealed a common trend of increased CNV burden in breast cancer cases, the trend appears to be strongest when assessing CNVs that overlap gene regions [73,74]. Evaluating such genes further by pathway analysis suggests two networks centred on factors known, TP53 and β-estradial [73], may be

important in breast cancer risk and development; however, these findings are yet to be reproduced. The feature of "CNV burden" has also been observed in the genome of patients with other cancers, suggesting that an uncharacterised subset of these variants may be causal [77–80]. Further studies are needed to identify recurring variants at shared loci.

5.3. Is There a Relationship between Germline CNVs and Breast Tumourigenesis

A characteristic of sporadic and familial breast tumours is genomic instability, resulting from either inherited mutations in genes that control genome integrity, or mutations that are acquired in somatic cells during development. Breast tumour cells in carriers of the *APOBEC3A-APOBEC3B* germline deletion show a greater number of C>T transitions than in non-carriers [81], thereby highlighting the importance of this common CNV in breast cancer development. It has previously been proposed that germline CNVs may also contribute to somatically acquired chromosome changes in tumours. Previous studies of Li-Fraumeni Syndrome (LFS) tumours [80] and of colon cancer-affected individuals [82] suggested that constitutional CNVs may act as a foundation on which chromosome copy number aberrations develop in tumour cells. These findings suggested a direct relationship between constitutional genomic variation and tumour genome evolution. The notion that inherited CNVs may influence the occurrence of somatically acquired copy number changes during breast cancer progression has not only prognostic significance, but also important consequences for early decisions relating to clinical management. Subsequent analyses of constitutional and tumour-specific CNVs in matched breast tumour and normal tissue using data from the Illumina Human CNV370 duo beadarray provided evidence that the location of copy number aberrations in tumour cells do not associate with constitutional CNVs [83]. However, the SNP arrays used in these studies had a relatively low number of probes and therefore poor spatial resolution for detecting CNVs and defining the variant boundaries. To determine the relationship between inherited genomic variation and genome evolution in breast cancer, sequencing-based studies are necessary to ensure accurate mapping of CNV breakpoints.

6. Conclusion

Genotyping constitutional CNVs using low- and high-resolution SNP arrays has served as the primary screening method for identifying potential genetic markers associated with breast cancer risk. Despite the large amount of SNP array data available from breast cancer studies, the contribution of inherited copy number variation to breast cancer risk remains relatively understudied. A variety of algorithms have been generated and matched to these datasets for predicting copy number-affected regions throughout the genome. Applying such algorithms may reveal new common and rare variants that contribute to breast cancer risk. However, initial analyses suggest array-based CNV data may be unreliable without further validation using ancillary technologies, such as qPCR, Nanostring, and MLPA. Moreover, the current and future use of new higher resolution technologies, including next-generation sequencing, will be critical for characterising CNV breakpoints, to better interpret their potential impact on breast cancer risk.

Acknowledgments

Author Contributions

Logan C. Walker conceived the review. Logan C. Walker, George A.R. Wiggins and John F. Pearson analysed literature, drafted and proofread the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. MacDonald, J.R.; Ziman, R.; Yuen, R.K.; Feuk, L.; Scherer, S.W. The database of genomic variants: A curated collection of structural variation in the human genome. *Nucleic Acids Res.* **2014**, *42*, D986–D992.
2. Zarrei, M.; MacDonald, J.R.; Merico, D.; Scherer, S.W. A copy number variation map of the human genome. *Nat. Rev. Genet.* **2015**, *16*, 172–183.
3. Database of Genomic Variants: A curated catalogue of human genomic structural variation. Available online: http://dgv.tcag.ca/dgv/app/home (accessed on 1 July 2015).
4. Stranger, B.E.; Forrest, M.S.; Dunning, M.; Ingle, C.E.; Beazley, C.; Thorne, N.; Redon, R.; Bird, C.P.; de Grassi, A.; Lee, C.; *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **2007**, *315*, 848–853.
5. Redon, R.; Ishikawa, S.; Fitch, K.R.; Feuk, L.; Perry, G.H.; Andrews, T.D.; Fiegler, H.; Shapero, M.H.; Carson, A.R.; Chen, W.; *et al.* Global variation in copy number in the human genome. *Nature* **2006**, *444*, 444–454.
6. Long, J.; Delahanty, R.J.; Li, G.; Gao, Y.T.; Lu, W.; Cai, Q.; Xiang, Y.B.; Li, C.; Ji, B.T.; Zheng, Y.; *et al.* A common deletion in the *APOBEC3* genes and breast cancer risk. *J. Natl. Cancer Inst.* **2013**, *105*, 573–579.
7. Girirajan, S.; Campbell, C.D.; Eichler, E.E. Human copy number variation and complex genetic disease. *Annu. Rev. Genet.* **2011**, *45*, 203–226.
8. Krepischi, A.C.; Pearson, P.L.; Rosenberg, C. Germline copy number variations and cancer predisposition. *Future Oncol.* **2012**, *8*, 441–450.
9. Palma, M.D.; Domchek, S.M.; Stopfer, J.; Erlichman, J.; Siegfried, J.D.; Tigges-Cardwell, J.; Mason, B.A.; Rebbeck, T.R.; Nathanson, K.L. The relative contribution of point mutations and genomic rearrangements in BRCA1 and BRCA2 in high-risk breast cancer families. *Cancer Res.* **2008**, *68*, 7006–7014.
10. Ziogas, A.; Gildea, M.; Cohen, P.; Bringman, D.; Taylor, T.H.; Seminara, D.; Barker, D.; Casey, G.; Haile, R.; Liao, S.Y.; *et al.* Cancer risk estimates for family members of a population-based family registry for breast and ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* **2000**, *9*, 103–111.

11. Hollestelle, A.; Wasielewski, M.; Martens, J.W.; Schutte, M. Discovering moderate-risk breast cancer susceptibility genes. *Curr. Opin. Genet. Dev.* **2010**, *20*, 268–276.

12. Walsh, T.; Casadei, S.; Coats, K.H.; Swisher, E.; Stray, S.M.; Higgins, J.; Roach, K.C.; Mandell, J.; Lee, M.K.; Ciernikova, S.; *et al.* Spectrum of mutations in *BRCA1*, *BRCA2*, *CHEK2*, and *TP53* in families at high risk of breast cancer. *JAMA* **2006**, *295*, 1379–1388.

13. Renwick, A.; Thompson, D.; Seal, S.; Kelly, P.; Chagtai, T.; Ahmed, M.; North, B.; Jayatilake, H.; Barfoot, R.; Spanova, K.; *et al.* ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat. Genet.* **2006**, *38*, 873–875.

14. Easton, D.F.; Pharoah, P.D.; Antoniou, A.C.; Tischkowitz, M.; Tavtigian, S.V.; Nathanson, K.L.; Devilee, P.; Meindl, A.; Couch, F.J.; Southey, M.; *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N. Engl. J. Med.* **2015**, *372*, 2243–2257.

15. Ahmed, S.; Thomas, G.; Ghoussaini, M.; Healey, C.S.; Humphreys, M.K.; Platte, R.; Morrison, J.; Maranian, M.; Pooley, K.A.; Luben, R.; *et al.* Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat. Genet.* **2009**, *41*, 585–590.

16. Antoniou, A.C.; Wang, X.; Fredericksen, Z.S.; McGuffog, L.; Tarrell, R.; Sinilnikova, O.M.; Healey, S.; Morrison, J.; Kartsonaki, C.; Lesnick, T.; *et al.* A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat. Genet.* **2010**, *42*, 885–892.

17. Cai, Q.; Long, J.; Lu, W.; Qu, S.; Wen, W.; Kang, D.; Lee, J.Y.; Chen, K.; Shen, H.; Shen, C.Y.; *et al.* Genome-wide association study identifies breast cancer risk variant at 10q21.2: Results from the Asia breast cancer consortium. *Hum. Mol. Genet.* **2011**, *20*, 4991–4999.

18. Easton, D.F.; Pooley, K.A.; Dunning, A.M.; Pharoah, P.D.; Thompson, D.; Ballinger, D.G.; Struewing, J.P.; Morrison, J.; Field, H.; Luben, R.; *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **2007**, *447*, 1087–1093.

19. Elgazzar, S.; Zembutsu, H.; Takahashi, A.; Kubo, M.; Aki, F.; Hirata, K.; Takatsuka, Y.; Okazaki, M.; Ohsumi, S.; Yamakawa, T.; *et al.* A genome-wide association study identifies a genetic variant in the SIAH2 locus associated with hormonal receptor-positive breast cancer in Japanese. *J. Hum. Genet.* **2012**, *57*, 766–771.

20. Fletcher, O.; Johnson, N.; Orr, N.; Hosking, F.J.; Gibson, L.J.; Walker, K.; Zelenika, D.; Gut, I.; Heath, S.; Palles, C.; *et al.* Novel breast cancer susceptibility locus at 9q31.2: Results of a genome-wide association study. *J. Natl. Cancer Inst.* **2011**, *103*, 425–435.

21. Garcia-Closas, M.; Couch, F.J.; Lindstrom, S.; Michailidou, K.; Schmidt, M.K.; Brook, M.N.; Orr, N.; Rhie, S.K.; Riboli, E.; Feigelson, H.S.; *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat. Genet.* **2013**, *45*, 392–398, e391–e392.

22. Gold, B.; Kirchhoff, T.; Stefanov, S.; Lautenberger, J.; Viale, A.; Garber, J.; Friedman, E.; Narod, S.; Olshen, A.B.; Gregersen, P.; *et al.* Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc. Natl. Acad. Sci. USA.* **2008**, *105*, 4340–4345.

23. Kim, H.C.; Lee, J.Y.; Sung, H.; Choi, J.Y.; Park, S.K.; Lee, K.M.; Kim, Y.J.; Go, M.J.; Li, L.; Cho, Y.S.; *et al.* A genome-wide association study identifies a breast cancer risk variant in *ERBB4* at 2q34: Results from the Seoul breast cancer study. *Breast Cancer Res.* **2012**, *14*, doi:10.1186/bcr3158.

24. Long, J.; Cai, Q.; Shu, X.O.; Qu, S.; Li, C.; Zheng, Y.; Gu, K.; Wang, W.; Xiang, Y.B.; Cheng, J.; *et al.* Identification of a functional genetic variant at 16q12.1 for breast cancer risk: Results from the Asia breast cancer consortium. *PLoS Genet.* **2010**, *6*, e1001002.

25. Long, J.; Cai, Q.; Sung, H.; Shi, J.; Zhang, B.; Choi, J.Y.; Wen, W.; Delahanty, R.J.; Lu, W.; Gao, Y.T.; *et al.* Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet.* **2012**, *8*, e1002532.

26. Low, S.K.; Takahashi, A.; Ashikawa, K.; Inazawa, J.; Miki, Y.; Kubo, M.; Nakamura, Y.; Katagiri, T. Genome-wide association study of breast cancer in the Japanese population. *PLoS ONE* **2013**, *8*, e76463.

27. Michailidou, K.; Beesley, J.; Lindstrom, S.; Canisius, S.; Dennis, J.; Lush, M.J.; Maranian, M.J.; Bolla, M.K.; Wang, Q.; Shah, M.; *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat. Genet.* **2015**, *47*, 373–380.

28. Michailidou, K.; Hall, P.; Gonzalez-Neira, A.; Ghoussaini, M.; Dennis, J.; Milne, R.L.; Schmidt, M.K.; Chang-Claude, J.; Bojesen, S.E.; Bolla, M.K.; *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **2013**, *45*, 353–361.

29. Stacey, S.N.; Manolescu, A.; Sulem, P.; Rafnar, T.; Gudmundsson, J.; Gudjonsson, S.A.; Masson, G.; Jakobsdottir, M.; Thorlacius, S.; Helgason, A.; *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* **2007**, *39*, 865–869.

30. Stacey, S.N.; Manolescu, A.; Sulem, P.; Thorlacius, S.; Gudjonsson, S.A.; Jonsson, G.F.; Jakobsdottir, M.; Bergthorsson, J.T.; Gudmundsson, J.; Aben, K.K.; *et al.* Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* **2008**, *40*, 703–706.

31. Thomas, G.; Jacobs, K.B.; Kraft, P.; Yeager, M.; Wacholder, S.; Cox, D.G.; Hankinson, S.E.; Hutchinson, A.; Wang, Z.; Yu, K.; *et al.* A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat. Genet.* **2009**, *41*, 579–584.

32. Turnbull, C.; Ahmed, S.; Morrison, J.; Pernet, D.; Renwick, A.; Maranian, M.; Seal, S.; Ghoussaini, M.; Hines, S.; Healey, C.S.; *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **2010**, *42*, 504–507.

33. Zheng, W.; Long, J.; Gao, Y.T.; Li, C.; Zheng, Y.; Xiang, Y.B.; Wen, W.; Levy, S.; Deming, S.L.; Haines, J.L.; *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.* **2009**, *41*, 324–328.

34. Peto, J.; Mack, T.M. High constant incidence in twins and other relatives of women with breast cancer. *Nat. Genet.* **2000**, *26*, 411–414.

35. Lin, P.; Hartz, S.M.; Wang, J.C.; Krueger, R.F.; Foroud, T.M.; Edenberg, H.J.; Nurnberger, J.I., Jr.; Brooks, A.I.; Tischfield, J.A.; Almasy, L.; *et al.* Copy number variation accuracy in genome-wide association studies. *Hum. Hered.* **2011**, *71*, 141–147.

36. Seiser, E.L.; Innocenti, F. Hidden markov model-based CNV detection algorithms for illumina genotyping microarrays. *Cancer Inform.* **2014**, *13*, 77–83.

37. Winchester, L.; Yau, C.; Ragoussis, J. Comparing CNV detection methods for SNP arrays. *Brief Funct. Genomic Proteomic* **2009**, *8*, 353–366.

38. Zhang, D.; Qian, Y.; Akula, N.; Alliey-Rodriguez, N.; Tang, J.; Gershon, E.S.; Liu, C. Accuracy of CNV detection from GWAS data. *PLoS ONE* **2011**, *6*, e14511.

39. Conrad, D.F.; Pinto, D.; Redon, R.; Feuk, L.; Gokcumen, O.; Zhang, Y.; Aerts, J.; Andrews, T.D.; Barnes, C.; Campbell, P.; *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **2010**, *464*, 704–712.

40. Kidd, J.M.; Cooper, G.M.; Donahue, W.F.; Hayden, H.S.; Sampas, N.; Graves, T.; Hansen, N.; Teague, B.; Alkan, C.; Antonacci, F.; *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **2008**, *453*, 56–64.

41. Komura, D.; Shen, F.; Ishikawa, S.; Fitch, K.R.; Chen, W.; Zhang, J.; Liu, G.; Ihara, S.; Nakamura, H.; Hurles, M.E.; *et al.* Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res.* **2006**, *16*, 1575–1584.

42. Marenne, G.; Rodriguez-Santiago, B.; Closas, M.G.; Perez-Jurado, L.; Rothman, N.; Rico, D.; Pita, G.; Pisano, D.G.; Kogevinas, M.; Silverman, D.T.; *et al.* Assessment of copy number variation using the Illumina Infinium 1M SNP-array: A comparison of methodological approaches in the Spanish Bladder Cancer/EPICURO study. *Hum. Mutat.* **2011**, *32*, 240–248.

43. Wang, K.; Li, M.; Hadley, D.; Liu, R.; Glessner, J.; Grant, S.F.; Hakonarson, H.; Bucan, M. PennCNV: An integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **2007**, *17*, 1665–1674.

44. Korn, J.M.; Kuruvilla, F.G.; McCarroll, S.A.; Wysoker, A.; Nemesh, J.; Cawley, S.; Hubbell, E.; Veitch, J.; Collins, P.J.; Darvishi, K.; *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVS. *Nat. Genet.* **2008**, *40*, 1253–1260.

45. Colella, S.; Yau, C.; Taylor, J.M.; Mirza, G.; Butler, H.; Clouston, P.; Bassett, A.S.; Seller, A.; Holmes, C.C.; Ragoussis, J. QuantiSNP: An objective Bayes hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* **2007**, *35*, 2013–2025.

46. Fiegler, H.; Redon, R.; Andrews, D.; Scott, C.; Andrews, R.; Carder, C.; Clark, R.; Dovey, O.; Ellis, P.; Feuk, L.; *et al.* Accurate and reliable high-throughput detection of copy number variation in the human genome. *Genome Res.* **2006**, *16*, 1566–1574.

47. van de Wiel, M.A.; Kim, K.I.; Vosse, S.J.; van Wieringen, W.N.; Wilting, S.M.; Ylstra, B. CGHcall: Calling aberrations for array CGH tumor profiles. *Bioinformatics* **2007**, *23*, 892–894.

48. Sun, W.; Wright, F.A.; Tang, Z.; Nordgard, S.H.; Van Loo, P.; Yu, T.; Kristensen, V.N.; Perou, C.M. Integrated study of copy number states and genotype calls using high-density SNP arrays. *Nucleic Acids Res.* **2009**, *37*, 5365–5377.

49. Price, T.S.; Regan, R.; Mott, R.; Hedman, A.; Honey, B.; Daniels, R.J.; Smith, L.; Greenfield, A.; Tiganescu, A.; Buckle, V.; *et al.* SW-ARRAY: A dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data. *Nucleic Acids Res.* **2005**, *33*, 3455–3464.

50. Day, N.; Hemmaplardh, A.; Thurman, R.E.; Stamatoyannopoulos, J.A.; Noble, W.S. Unsupervised segmentation of continuous genomic data. *Bioinformatics* **2007**, *23*, 1424–1426.

51. Scharpf, R.B.; Parmigiani, G.; Pevsner, J.; Ruczinski, I. Hidden Markov models for the assessment of chromosomal alterations using high-throughput SNP arrays. *Ann. Appl. Stat.* **2008**, *2*, 687–713.

52. Coin, L.J.; Asher, J.E.; Walters, R.G.; Moustafa, J.S.; de Smith, A.J.; Sladek, R.; Balding, D.J.; Froguel, P.; Blakemore, A.I. cnvHap: An integrative population and haplotype-based multiplatform model of SNPs and CNVs. *Nat. Methods* **2010**, *7*, 541–546.

53. Li, C.; Beroukhim, R.; Weir, B.A.; Winckler, W.; Garraway, L.A.; Sellers, W.R.; Meyerson, M. Major copy proportion analysis of tumor samples using SNP arrays. *BMC Bioinformatics* **2008**, *9*, doi:10.1186/1471-2105-9-204.

54. Pique-Regi, R.; Caceres, A.; Gonzalez, J.R. R-gada: A fast and flexible pipeline for copy number analysis in association studies. *BMC Bioinformatics* **2010**, *11*, doi:10.1186/1471-2105-11-380.

55. Gai, X.; Perin, J.C.; Murphy, K.; O'Hara, R.; D'Arcy, M.; Wenocur, A.; Xie, H.M.; Rappaport, E.F.; Shaikh, T.H.; White, P.S. CNV workshop: An integrated platform for high-throughput copy number variation discovery and clinical diagnostics. *BMC Bioinformatics* **2010**, *11*, doi:10.1186/1471-2105-11-74.

56. Korbel, J.O.; Urban, A.E.; Affourtit, J.P.; Godwin, B.; Grubert, F.; Simons, J.F.; Kim, P.M.; Palejev, D.; Carriero, N.J.; Du, L.; *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **2007**, *318*, 420–426.

57. Eckel-Passow, J.E.; Atkinson, E.J.; Maharjan, S.; Kardia, S.L.; de Andrade, M. Software comparison for evaluating genomic copy number variation for Affymetrix 6.0 SNP array platform. *BMC Bioinformatics* **2011**, *12*, doi:10.1186/1471-2105-12-220.

58. Pinto, D.; Darvishi, K.; Shi, X.; Rajan, D.; Rigler, D.; Fitzgerald, T.; Lionel, A.C.; Thiruvahindrapuram, B.; Macdonald, J.R.; Mills, R.; *et al.* Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat. Biotechnol.* **2011**, *29*, 512–520.

59. Dupuis, M.C.; Zhang, Z.; Durkin, K.; Charlier, C.; Lekeux, P.; Georges, M. Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. *Anim. Genet.* **2013**, *44*, 206–208.

60. Metzger, J.; Philipp, U.; Lopes, M.S.; da Camara Machado, A.; Felicetti, M.; Silvestrelli, M.; Distl, O. Analysis of copy number variants by three detection algorithms and their association with body size in horses. *BMC Genomics* **2013**, *14*, doi:10.1186/1471-2164-14-487.

61. Lin, Y.J.; Chen, Y.T.; Hsu, S.N.; Peng, C.H.; Tang, C.Y.; Yen, T.C.; Hsieh, W.P. HaplotypeCN: Copy number haplotype inference with hidden Markov model and localized haplotype clustering. *PLoS ONE* **2014**, *9*, e96841.

62. Park, H.; Kim, J.I.; Ju, Y.S.; Gokcumen, O.; Mills, R.E.; Kim, S.; Lee, S.; Suh, D.; Hong, D.; Kang, H.P.; *et al.* Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. *Nat. Genet.* **2010**, *42*, 400–405.

63. Zhang, X.; Du, R.; Li, S.; Zhang, F.; Jin, L.; Wang, H. Evaluation of copy number variation detection for a SNP array platform. *BMC Bioinformatics* **2014**, *15*, doi:10.1186/1471-2105-15-50.

64. Komatsu, A.; Nagasaki, K.; Fujimori, M.; Amano, J.; Miki, Y. Identification of novel deletion polymorphisms in breast cancer. *Int. J. Oncol.* **2008**, *33*, 261–270.

65. Erikson, G.A.; Deshpande, N.; Kesavan, B.G.; Torkamani, A. SG-ADVISER CNV: Copy-number variant annotation and interpretation. *Genet Med* **2014**, *8*, doi:10.1038/gim.2014.180.

66. Vandeweyer, G.; Reyniers, E.; Wuyts, W.; Rooms, L.; Kooy, R.F. CNV-webstore: Online CNV analysis, storage and interpretation. *BMC Bioinformatics* **2011**, *12*, doi:10.1186/1471-2105-12-4.

67. Zhao, M.; Zhao, Z. CNVannotator: A comprehensive annotation server for copy number variation in the human genome. *PLoS ONE* **2013**, *8*, e80170.

68. Online Mendelian Inheritance in Man: An online catalog of human genes and genetic disorders. Available online: http://www.omim.org/ (accessed on 1 July 2015).

69. ClinVar. Available online: www.ncbi.nlm.nih.gov/clinvar (accessed on 1 July 2015).

70. Behjati, S.; Huch, M.; van Boxtel, R.; Karthaus, W.; Wedge, D.C.; Tamuri, A.U.; Martincorena, I.; Petljak, M.; Alexandrov, L.B.; Gundem, G.; *et al.* Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **2014**, *513*, 422–425.

71. Xuan, D.; Li, G.; Cai, Q.; Deming-Halverson, S.; Shrubsole, M.J.; Shu, X.O.; Kelley, M.C.; Zheng, W.; Long, J. APOBEC3 deletion polymorphism is associated with breast cancer risk among women of european ancestry. *Carcinogenesis* **2013**, *34*, 2240–2243.

72. Handsaker, R.E.; Van Doren, V.; Berman, J.R.; Genovese, G.; Kashin, S.; Boettger, L.M.; McCarroll, S.A. Large multiallelic copy number variations in humans. *Nat. Genet.* **2015**, *47*, 296–303.

73. Pylkas, K.; Vuorela, M.; Otsukka, M.; Kallioniemi, A.; Jukkola-Vuorinen, A.; Winqvist, R. Rare copy number variants observed in hereditary breast cancer cases disrupt genes in estrogen signaling and TP53 tumor suppression network. *PLoS Genet.* **2012**, *8*, e1002734.

74. Kuusisto, K.M.; Akinrinade, O.; Vihinen, M.; Kankuri-Tammilehto, M.; Laasanen, S.L.; Schleutker, J. Copy number variation analysis in familial BRCA1/2-negative Finnish breast and ovarian cancer. *PLoS ONE* **2013**, *8*, e71802.

75. Jakobsson, M.; Scholz, S.W.; Scheet, P.; Gibbs, J.R.; VanLiere, J.M.; Fung, H.C.; Szpiech, Z.A.; Degnan, J.H.; Wang, K.; Guerreiro, R.; *et al.* Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* **2008**, *451*, 998–1003.

76. Krepischi, A.C.; Achatz, M.I.; Santos, E.M.; Costa, S.S.; Lisboa, B.C.; Brentani, H.; Santos, T.M.; Goncalves, A.; Nobrega, A.F.; Pearson, P.L.; *et al.* Germline DNA copy number variation in familial and early-onset breast cancer. *Breast Cancer Res.* **2012**, *14*, doi:10.1186/bcr3109.

77. Moir-Meyer, G.L.; Pearson, J.F.; Lose, F.; Scott, R.J.; McEvoy, M.; Attia, J.; Holliday, E.G.; Pharoah, P.D.; Dunning, A.M.; Thompson, D.J.; *et al.* Rare germline copy number deletions of likely functional importance are implicated in endometrial cancer predisposition. *Hum. Genet.* **2014**, *134*, 269–278.

78. Talseth-Palmer, B.A.; Holliday, E.G.; Evans, T.J.; McEvoy, M.; Attia, J.; Grice, D.M.; Masson, A.L.; Meldrum, C.; Spigelman, A.; Scott, R.J. Continuing difficulties in interpreting cnv data: Lessons from a genome-wide CNV association study of Australian HNPCC/lynch syndrome patients. *BMC Med. Genomics* **2013**, *6*, doi:10.1186/1755-8794-6-10.

79. Yang, R.; Chen, B.; Pfutze, K.; Buch, S.; Steinke, V.; Holinski-Feder, E.; Stocker, S.; von Schonfels, W.; Becker, T.; Schackert, H.K.; *et al.* Genome-wide analysis associates familial colorectal cancer with increases in copy number variations and a rare structural variation at 12p12.3. *Carcinogenesis* **2014**, *35*, 315–323.

80. Shlien, A.; Tabori, U.; Marshall, C.R.; Pienkowska, M.; Feuk, L.; Novokmet, A.; Nanda, S.; Druker, H.; Scherer, S.W.; Malkin, D. Excessive genomic DNA copy number variation in the Li-Fraumeni cancer predisposition syndrome. *Proc. Natl. Acad. Sci. USA.* **2008**, *105*, 11264–11269.

81. Nik-Zainal, S.; Wedge, D.C.; Alexandrov, L.B.; Petljak, M.; Butler, A.P.; Bolli, N.; Davies, H.R.; Knappskog, S.; Martin, S.; Papaemmanuil, E.; *et al.* Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat. Genet.* **2014**, *46*, 487–491.

82. Camps, J.; Grade, M.; Nguyen, Q.T.; Hormann, P.; Becker, S.; Hummon, A.B.; Rodriguez, V.; Chandrasekharappa, S.; Chen, Y.; Difilippantonio, M.J.; *et al.* Chromosomal breakpoints in primary colon cancer cluster at sites of structural variants in the genome. *Cancer Res.* **2008**, *68*, 1284–1295.

83. Walker, L.C.; Krause, L.; Spurdle, A.B.; Waddell, N. Germline copy number variants are not associated with globally acquired copy number changes in familial breast tumours. *Breast Cancer Res. Treat* **2012**, *134*, 1005–1011.